<u>IN THE SPECIFICATION</u>:

Please amend paragraph [0002] starting at page 1, line 14 as follows.

--Recently, speech synthesis apparatuses are known in which speech synthesis is performed by inputting text data, generating prosody parameters while generating silent positions, the lengths of silent times, accents and the like by performing language analysis of the text data, <u>and</u> retrieving a synthesis units inventory storing synthesis units in accordance with the prosody parameters.--

Please amend paragraph [0003] starting at page 1, line 20 as follows.

--Such speech synthesis apparatuses mainly adopt a PSOLA (pitch-synchronous overlap-add) method in which the retrieved units are ~~modified~~ <u>modified,</u> by copying or deleting each pitch ~~waveforms~~ <u>waveform</u> consisting of the units, and concatenated <u>with</u> each other.--

Please amend paragraph [0027] starting at page 6, line 21 as follows.

--The D/A converter 105 converts speech-waveform data (a digital signal) formed by executing the control program into an analog signal, and outputs the analog signal to the speaker 109. Even if the speaker 109 is not provided in the main body of the apparatus, it is also possible to output the analog signal from a speaker of another apparatus via a network. In this case, an analog signal obtained by converting a digital signal by the D/A converter 105 may be output to another terminal via the network. Alternatively, it is, of course, possible to output a digital signal to another terminal via a network, ~~converted~~ <u>convert</u> the digital signal into an

2

analog signal at the terminal, and output the analog signal. Particularly when outputting an analog signal, a terminal where the analog signal is output via a network may include only a speaker. Hence, the terminal is not limited to a computer, but may be a telephone set, a portable terminal or an audio apparatus. Even such a terminal can deal with a case of receiving a digital signal if a D/A converter is included.--

Please amend paragraph [0043] starting at page 12, line 9 as follows.

--In this equation, a weighting coefficient w is empirically obtained by ~~an~~ a preliminary experiment or the like. In the case of w = 0, distortion values are described only by modification distortions Dm. In the case of w = 1, distortion values depend only on concatenation distortions Dc.--

Please amend paragraph [0045] starting at page 12, line 21 as follows.

--FIG. 6 is a diagram illustrating how to obtain a ~~connection~~ concatenation distortion Dc in the first embodiment.--

Please amend paragraph [0049] starting at page 14, line 25 as follows.

--As described above, according to the first embodiment, by performing speech synthesis by obtaining a concatenation distortion and a ~~modulation~~ modification distortion for each synthesis unit, obtaining a distortion value of each synthesis unit by performing weighting

calculation based on the obtained distortions, and specifying a synthesis-unit series having a minimum sum of distortion values, it is possible to obtain an excellent result of speech synthesis.--

Please amend paragraph [0050] starting at page 15, line 7 as follows.

--Although in the first embodiment, the case of using a diphone as a synthesis unit <u>has been described</u>, the present invention is not limited to such an approach. For example, a phoneme or a half-diphone may be adopted as a synthesis unit. The half-diphone is obtained by dividing a diphone into two portions at a border of phonemes.--

Please amend paragraph [0051] starting at page 15, line 12 as follows.

--FIG. 8 is a schematic diagram when the half-diphone is used as a unit. Merits in such an approach will now be briefly described. When synthesizing an arbitrary text, a synthesis units inventory must prepare all types of diphones. On the other hand, when using the half-diphone as a unit, a half-diphone which ~~lacks~~ <u>is lacking</u> can be substituted by another half-diphone. For example, even if "/a.n.0/" is used instead of ~~"/a.b.0/(the~~ "/a.b.0/ <u>(the</u> left side of a diphone a.b)", a voice can be excellently reproduced with less degradation of quality. Hence, the size of the synthesis units inventory can be reduced.--

Please amend paragraph [0058] starting at page 17, line 7 as follows.

--Although in the first embodiment, the case of using cepstrum for calculating a concatenation distortion <u>has been described</u>, the present invention is not limited to such an

4

approach. For example, a concatenation distortion may be obtained using the sum of differences between waveforms before and after a concatenation point. Alternatively, a concatenation distortion may be obtained using, for example, a spectrum distance. In this case, a concatenation point is preferably synchronized with a pitch mark.--

Please amend paragraph [0060] starting at page 17, line 22 as follows.

--Although in the first embodiment, the case of using the sum of differences for each order of cepstrum for calculation of a concatenation distortion has been described, the present invention is not limited to such an approach. For example, each order may be normalized (normalization coefficient rj) using statistical properties or the like. In this case, a concatenation distortion Dc is expressed by:

$$Dc = \Sigma \Sigma (rj \times |Cprei,j - Ccur\ i,j|),$$

where the first Σ indicates the sum of the case in which i changes from -2 to 2, and the second Σ indicates the sum of the case in which j changes from 0 to 16.--

Please amend paragraph [0063] starting at page 18, line 26 as follows.

--Although in the first embodiment, the case of calculating a modification distortion based on information obtained from a waveform has been described, the present invention is not limited to such an approach. For example, A a modification distortion may be calculated based on the number of operations of deleting and copying a pitch waveform unit when performing a PSOLA operation.--

5

Please amend paragraph [0066] starting at page 19, line 23 as follows.

--Although in the first embodiment, the case of calculating a modification distortion every time a synthesis unit is modified during speech synthesis has been described, the present invention is ~~no~~ <u>not</u> limited to such an approach. For example, modification distortions may be calculated in advance and stored in a table.--